This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

Proc. Natl. Acad. Sci. USA
Vol. 79, pp. 6598–6602, November 1982
Genetics

Open reading frame cloning: Identification, cloning, and expression of open reading frame DNA

(*lacZ* gene fusion/frameshift mutant/hybrid protein)

MARK R. GRAY*, HILDUR V. COLOT*, LEONARD GUARENTE†, AND MICHAEL ROSBASH*

*Department of Biology, Brandeis University, Waltham, Massachusetts 02254; and †Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Communicated by Norman Davidson, July 14, 1982

ABSTRACT **A plasmid was constructed that facilitates the cloning and expression of open reading frame DNA. A DNA fragment containing a bacterial promoter and the amino terminus of the *cI* gene of bacteriophage λ was fused to an amino-terminally deleted version of the *lacZ* gene. An appropriate cloning site was inserted between these two fragments such that a frameshift mutation was introduced upstream of the lacZ-encoding DNA. This cloning vehicle produces a relatively low level of β-galactosidase activity when introduced into *Escherichia coli.* The insertion of foreign DNA at the cloning site can reverse the frameshift mutation and generate plasmids that produce a relatively high level of β-galactosidase activity. A large fraction of these plasmids produce a fusion protein that has a portion of the λ cI protein at the amino terminus, the foreign protein segment in the middle, and the lacZ polypeptide at the carboxyl terminus. The production of a high level of β-galactosidase and a large fusion polypeptide guarantees the cloning of a DNA fragment with at least one open reading frame that traverses the entirety of the fragment. Hence, the method can identify, clone, and express (as part of a larger fusion polypeptide) open reading frame DNA from among a large collection of DNA fragments.**

The use of recombinant DNA technology has increased our understanding of the organization and expression of eukaryotic DNA. While many goals can be achieved with existing technology, certain features of eukaryotic gene organization are still attainable only with difficulty. In particular, the identification of protein-coding segments within a large block of eukaryotic DNA is quite challenging. One should be able to take advantage of properties that are unique or highly enriched in coding DNA. Perhaps the most ubiquitous and specific feature of coding DNA is the presence of open reading frames. Indeed, analysis at the DNA sequence level has proved invaluable in identifying genes or gene segments. Although there are many examples of genes in which relatively small exons are interrupted by relatively large and numerous introns, this characteristic appears to be relatively infrequent in lower eukaryotes. Also, higher eukaryotic genes have been analyzed that contain open reading frames of substantial length.

In this report, we describe a method for cloning pieces of DNA based on this property. We have exploited the elegant work of Beckwith and co-workers (1), who have demonstrated the power and utility of gene and operon fusions to the *Escherichia coli* β-galactosidase gene (the *lacZ* gene). Many of these studies rely on the fact that the enzyme, β-galactosidase (β-D-galactoside galactohydrolase, EC 3.2.1.23) is often biologically active when it carries other protein segments at its amino terminus. Indeed, it appears that in yeast, as well as in *E. coli*, β-galactosidase can carry a wide range of protein sequences at its amino terminus and still retain substantial biological activity (2, 3). We have taken advantage of this property to create a *lacZ* vector for the cloning of open reading frames. We have introduced a frameshift mutation by inserting a DNA linker at the junction of a *cIlacZ* fused gene. As expected, this frameshift results in a relatively low level of *lacZ* activity. The DNA linker encodes a cloning site. Consequently, the vector is able to screen for segments of open reading frame DNA, since a continuous open reading frame sequence is a necessary condition for high-level expression of the *lacZ* sequence downstream from the insert. A successful construction results in the synthesis of a lacZ fusion polypeptide that includes the protein segment coded for by the inserted open reading frame. The vector is able to identify, clone, and express open reading frame DNA in one step.

METHODS AND MATERIALS

Materials. Restriction enzymes, T4 DNA ligase, and BAL-31 nuclease were purchased from New England BioLabs and Bethesda Research Laboratories. Calf intestine phosphatase was purchased from Boehringer-Mannheim; 5-bromo-4-chloro-3-indolyl β-D-galactoside (XGal), *o*-nitrophenyl β-D-galactopyranoside (ONPG), and isopropyl β-D-thiogalactoside were purchased from Sigma. XGal plates and MacConkey agar plates (Difco) were prepared as described by Miller (4). *Sma I/BamHI* adapter was purchased from New England BioLabs.

Antibodies. Rabbit anti-β-galactosidase was a gift of Beth Rasmussen, rabbit anti-cI was a gift of Jeffrey Roberts, and the ^{125}I-labeled IgG fraction of a goat anti-rabbit immunoglobulin Fc serum was a gift of Susan Lowey and Joan Press.

Bacterial Strains. The lac deletion strain LG90 (F$^-$ Δ*lac pro* XIII) was used exclusively (2).

Plasmid DNA Preparation. Plasmid DNA was prepared by standard CsCl/ethidium bromide centrifugation methods (5) or from minilysates prepared by using the alkaline lysis procedure (6).

Plasmid Constructions. Plasmid DNA was digested with commercial restriction enzymes as recommended by the suppliers. Where indicated, restricted DNA was treated with sufficient phosphatase [10 mM Tris·HCl (pH 8.0), 30 min at 37°C] to reduce self-ligation of a comparable amount of vector DNA by a factor of at least 1/100. DNA was sonicated in 5 ml of TE buffer (10 mM Tris·HCl, pH 8.3/1 mM EDTA)/0.2 M NaCl with six 30-sec bursts at maximum power, concentrated on a 0.2-ml column of DEAE-cellulose, eluted with TE buffer/1 M NaCl, and precipitated with ethanol. DNA was resuspended in TE buffer, digested lightly with BAL-31 nuclease (sufficient to

---

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Abbreviations: XGal, 5-bromo-4-chloro-3-indolyl β-D-galactoside; bp, base pair(s).

remove approximately 25 nucleotides from each end of a similar quantity of control restriction enzyme-digested DNA fragments) at 30°C, and fractionated by size on 10% acrylamide/Tris borate/EDTA (TBE) gels along with appropriate molecular weight markers (7).

Ligations. Cohesive-end ligations were carried out as follows: 100 ng of plasmid vector and various amounts of eukaryotic DNA were ligated for 3 hr at 22°C in a final volume of 100 μl with 0.1 (Bethesda Research Laboratories) unit of T4 DNA ligase. Blunt-end ligations were typically carried out as follows: 500 ng of plasmid vector and various amounts of potential insert DNA were ligated overnight at 16°C in a final volume of 10 μl with 1 unit of T4 DNA ligase.

Transformation. Bacterial transformation into LG90 was carried out by standard procedures (8), with the following modifications. All ligated DNA samples were phenol extracted and ethanol precipitated prior to transformation, which caused a reproducible 5- to 10-fold increase in transformation efficiency. Transformed cells (1 ml) were spread on the appropriate indicator plates containing ampicillin at 25 μg/ml. Colonies were scored for phenotype on MacConkey agar plates after 24 hr at 37°C. If the plates were crowded, the colony phenotype was scored at 48 hr.

Hybrid Protein Analysis. Three-milliliter samples of cells from a saturated culture (grown in L broth containing ampicillin at 50 μg/ml) were centrifuged, the pellet was suspended in 150 μl of 1.2-fold concentrated sample buffer (9) and heated at 100°C for 3 min, and the viscosity was reduced by repeated passages through a 22-gauge syringe needle. Between 4 and 8 μl of the protein samples were electrophoresed in 7.5% acrylamide gels (9). Protein blotting was carried out according to published procedures (10).

β-Galactosidase Assays. These were carried out according to Miller (4) except that cells were grown in L broth.

RESULTS

Our initial hypothesis was that a frameshift mutation placed near the amino terminus of an open reading frame sequence that contains the *lacZ* sequence at its carboxyl terminus would produce a plasmid that would be phenotypically lacZ$^-$ when transformed into *E. coli*. Furthermore, the insertion of foreign DNA at or near the frameshift could correct the frameshift mutation and generate a *lacZ*$^+$ insert DNA-containing plasmid. The two constructions shown in Fig. 1*A* were designed to test this hypothesis. A segment of the *cI* gene from bacteriophage λ terminating at a *Hin*dIII site at codon 157 was fused to a *lacIZ* fragment. The *cI* gene fragment also contains two *lac* promoters that are oriented toward *cI*. Both constructions were identical except that two different *lacZ*-containing plasmids were used. These two plasmids, pLG200 and pLG400, differ in DNA sequence only in the immediate vicinity of the *Hin*dIII site (12). Because the DNA sequence of the λ *cI* gene is also known (13), as well as the reading frame of the lacI–lacZ (lacIZ) fusion protein encoded by both pLG200 and pLG400 (12), we could predict the reading frame of the two resultant plasmids. Plasmid pMR1 has 24 bases between the lysine codon at amino acid 157 in the *cI* gene and the leucine codon at the beginning of the *lacI* fragment (Fig. 1*A*). Since none of these eight codons are stop codons, this construction reads in frame from the λ *cI* gene into the *lacIZ* fusion and consequently produces a high level of β-galactosidase activity. In contrast, pMR2, constructed from pLG200, has 10 bases between the AAG lysine codon of the λ *cI* gene and the CTG leucine codon of *lacI*. Because this number of intervening bases is not divisible by 3, this construction should shift the frame, resulting in improper reading of or termination in (or both) the *lacIZ* fusion sequence. This plasmid results in a relatively low level of β-galactosidase activity. In addition, the insertion of eukaryotic DNA into the *Hin*dIII site of pMR2 is able to generate colonies with a high level of lacZ activity (data not shown).

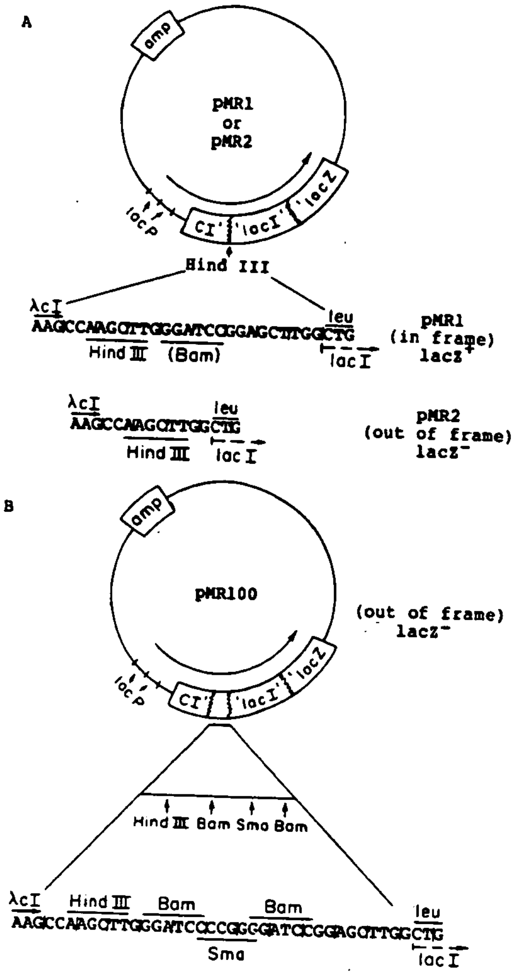

FIG. 1. Construction of vectors. (*A*) Plasmids pKB252 (11), pLG200, and pLG400 (12) were manipulated by standard procedures to produce the appropriate fragments for ligation. Briefly, pKB252 was digested with *Hae* III/*Hin*dIII and the largest fragment was purified. pLG200 (or pLG400) was digested with *Sma* I/*Hin*dIII and the largest fragment was purified. These two fragments were ligated together and ampicillin-resistant colonies were selected. The restriction sites of the resultant plasmids were verified. pMR1 was derived by using pLG400 and pMR2 was derived by using pLG200. Both plasmids, when transformed into LG90, form blue colonies on XGal plates. Thus, lac$^+$ and lac$^-$ refer to phenotypes on MacConkey agar plates. The DNA sequence at the joint is inferred from the sequence of the λ *cI* gene up to the *Hin*dIII site (13) and the sequence in the vicinity of the *Hin*dIII site of pLG200 and pLG400 (12). (*B*) One hundred nanograms of *Bam*HI-digested pMR1 and 2 ng of *Bam*HI/*Sma* I adapter were ligated (final vol, 100 μl) with 0.1 unit of ligase for 3 hr at room temperature. DNAs were transformed into LG90, and ampicillin-resistant lac$^-$ colonies were scored on MacConkey agar plates. DNA sequences in the vicinity of the cloning sites were verified by direct sequence analysis.

To construct a vector of more utility, pMR1 was manipulated as in Fig. 1*B*. The construction inserted 10 base pairs (bp) at the *Bam*HI site and therefore changed the reading frame so that the *lacZ* sequence is no longer properly translated. Moreover, it has created a blunt-ended *Sma* I site, flanked by two *Bam*HI sites, within this frameshifted region. The *Sma* I site provides a cloning site into which blunt-ended DNA can be cloned. Insert DNA can be excised by digestion with *Bam*HI.

To verify that this second-generation vector was suitable for cloning open reading frame DNA, *Hae* III-digested pBR322 was cloned into the *Sma* I site of pMR100. After transformation into LG90, 28 red colonies were chosen at random and analyzed for plasmid DNA insert size. All 28 clones yielded an additional *Bam*HI fragment when analyzed by acrylamide gel electrophoresis; 8 of these clones are shown in Fig. 2. A very limited subset of the possible pBR322 *Hae* III fragments are cloned when ampicillin-resistant colonies are chosen on the basis of a strong lacZ$^+$ phenotype. As expected, most white colonies from this ligation contain plasmid DNA with no detectable insert, since most of these colonies are presumably derived from recyclization of the vector (see Table 1). White colonies that harbor a plasmid with a DNA insert do not show the dramatic insert size preference seen with red colonies; DNA fragments of many different sizes are obtained from white colonies (data not shown).

Table 1. Cloning of pBR322 *Hae* III fragments into pMR100 and pMR200

Ligation mixture	Vector	Vector/pBR322 ratio*	Colonies	
			Red	White
1	pMR200†	1:1	349	13
2	pMR200	1:5	300	65
3	pMR100	1:5	11	594

Phosphatase-digested pBR322 *Hae* III fragments were ligated with *Sma* I-digested pMR100 or *Sma* I-digested pMR200. Ligated DNA was transformed into LG90 and colonies were scored on MacConkey agar plates. Because pMR200 is a lacZ$^+$ vector having the *Sma* I cloning site at the same position as pMR100, the fraction of white colonies ≈ the fraction of religated vector that contains an insert. This value was ≈20% in ligation mixture 2 but, in ligation mixture 3 under identical conditions, 11/605 colonies are red. Normalizing for the expected fraction of colonies from this ligation that contain an insert (≈20% ≈ same value as ligation mixture 2), the fraction of pBR322 *Hae* III fragments that can reverse the frameshift ≈ (11/605) × 0.2 ≈ 10%.

* Arbitrary.

† pMR200 was derived from pMR100 as a rare lacZ$^+$ colony (on MacConkey agar) after digestion of pMR100 with *Sma* I, self-ligation, and transformation. It is almost certainly a single-base-pair deletion of pMR100, placing the reading frame back in register. [The *Sma* I site (Fig. 1) is retained despite the deletion.] The size of the lacZ polypeptide is identical to that produced by pMR1 and all restriction sites in pMR100 are still present. Since almost all inserts into the *Sma* I site are lacZ$^-$, the vector can be used to measure insert frequency.

The sequence of pMR100 predicts that two criteria must be met for a DNA fragment inserted into the *Sma* I site to generate a proper open reading frame between the λ *cI* fragment and the *lacIZ* fragment. (*i*) The inserted DNA must be of a size 3*n* − 1 (where *n* is an integer) and (*ii*) the inserted DNA must contain no stop codons in the reading frame set by the frame of the λ *cI* gene. Examination of the size and sequence (in both orientations) of the pBR322 *Hae* III fragments predicts that, of the 22 *Hae* III fragments, only 3 meet the two criteria: a 104-bp fragment from the tetracycline-resistance gene (which contains the single pBR322 *Bam*HI site), an 89-bp fragment, and an 80-bp fragment, the latter two from the ampicillin-resistance gene. The 89-bp fragment has an appropriate open reading frame in both orientations, while the 104-bp fragment is cleaved by *Bam*HI into two fragments, a 78- and a 26-bp fragment. Consequently, the sequence data predict that 4 fragment orientations out of a possible 44 (22 fragments × two orientations) should produce red colonies. This value of 10% is consistent with experiment (Table 1). Moreover, there should be two size classes of insert fragments produced by *Bam*HI digestion, 103 and about 93 nucleotides (89- and 78- to 80-nucleotide fragments plus 14 nucleotides from the DNA between the *Bam*/*Sma* sites shown in Fig. 1). Consistent with expectation, 26 of the 28 red colonies examined at random have one of the two size classes of DNA insert shown in Fig. 2. Moreover, double digestion of plasmids derived from these red colonies confirms that some of the larger size DNA inserts are indeed derived from the 89-bp *Hae* III fragment, which has been cloned in both orientations.

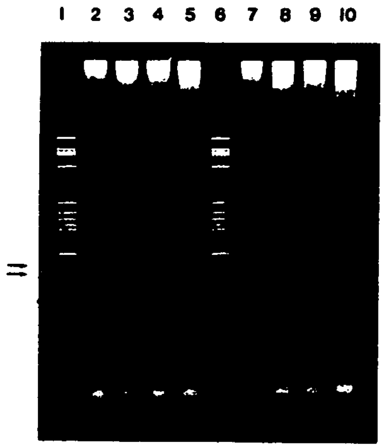

FIG. 2. Cloning of *Hae* III fragments of pBR322 into pMR100. pBR322 DNA was digested to completion with *Hae* III and the resultant fragments digested with phosphatase (to prevent multiple inserts). The phosphatase-digested DNA was ligated with pMR100 and transformed into LG90. Red colonies were picked, and DNA was isolated from minilysates and digested with *Bam*HI. DNA was analyzed on 10% acrylamide/Tris borate/EDTA gels. Lanes: 1 and 6, *Hae* III pBR322 fragment standards; 2–5 and 7–10, DNA from random red colonies. Arrows indicate the 104-bp and 89-bp *Hae* III pBR322 fragments.

The above data suggest that the vector pMR100 can select suitable open reading frame fragments from among a larger group of DNA segments. Because of the nature of the cloning vehicle, a productive DNA insert should be expressed as a part of a larger fusion polypeptide. To verify this prediction, we cloned into pMR100 sonicated DNA (carefully size fractionated to avoid cloning small DNA fragments) derived from a cloned piece of *Drosophila melanogaster* DNA. Eleven red colonies were chosen for detailed analysis; the data are summarized in Table 2. Five of the 11 colonies contain proteins that are bigger (by approximately 20,000 daltons) than the initial cI–lacIZ fusion protein, consistent with a DNA insert size of approximately 500 bp (Table 2 and Fig. 3). These large proteins react with both anti-cI and anti-β-galactosidase antisera (Table 2 and Fig. 4), consistent with the hypothesis that they are proteins of the general structure cI–eukaryotic piece–lacIZ. These five strains contain additional smaller proteins, the largest of which is approximately 20,000 daltons less than the initial cI–lacZ fusion protein of pMR200. This protein reacts positively with anti-β-galactosidase antiserum but does not react with anti-cI antiserum. A protein with similar properties is also visible in the strain with pMR200. Two of the colonies (nos. 2 and 4) have properties suggesting that they are lacZ$^+$ revertants of pMR100 generated by a deletion of perhaps 1 bp during cloning. Four of the colonies (nos. 7, 9, 10, and 11) contain only the smaller proteins

Table 2. Insertion of cloned sonicated *Drosophila* DNA into pMR100

	Insert size	β-Gal activity	Protein size*	Protein blot	
				anti-cI	anti-β-Gal
Clone					
1	425	1,464	Small, big†	Big	Small, big
2	—	ND	pMR200	ND	ND
3	515	1,547	Small, big	Big	Small, big
4	—	ND	pMR200	ND	ND
5	425	1,569	Small, big	Big	Small, big
6	520	1,885	Small, big	Big	Small, big
7	460	660	Small	—	Small
8	440	795	Small, big	Big	Small, big
9	400	762	Small	—	Small
10	530	441	Small	—	Small
11	400	351	Small	—	Small
LG90	—	2	—	ND	ND
pMR100	—	30	—	—	—
pMR200	—	10,661	pMR200	Big	Small, big

A *Drosophila melanogaster* clone, pDm2837, was a generous gift of Welcome Bender. This eukaryotic insert, an 8.1-kilobase (kb) *Sal* I fragment in the *Sal* I site of pBR322, has been identified on the basis of genetic criteria to include the *ry*$^+$ gene, which codes for xanthine dehydrogenase (Welcome Bender and Arthur Chovnick, personal communication). The 8.1-kb *Sal* I insert was purified by agarose gel electrophoresis and electroelution and then sonicated. DNAs of 400–550 bp were selected for cloning, and 11 red colonies were chosen for subsequent analysis. No colonies or data were discarded from the analysis. β-Gal, β-galactosidase; —, no detectable insert or protein band; ND, not done.

* Protein size refers to the prominent β-galactosidase polypeptide(s) visible on NaDodSO$_4$ gels as assayed by Coomassie blue staining and blotting with anti-β-galactosidase antibody.

† "Big" refers to a β-galactosidase polypeptide 15,000–20,000 daltons larger than the cI–lacIZ fusion in pMR200. "Small" refers to a β-galactosidase polypeptide 15,000–20,000 daltons smaller than the cI–lacIZ fusion in pMR200; none of these "small" proteins reacts with anti-cI-antibody.

(no. 7 is shown in Figs. 3 and 4). The properties of these polypeptides are identical to those of the small protein of the other clones described above. The origin of this small protein is uncertain (see *Discussion*).

DISCUSSION

The experiments presented here validate our initial hypothesis: a plasmid vector having a frameshift near the amino terminus of a gene that codes for a polypeptide with lacZ$^+$ activity can be used to identify, clone, and express open reading frame segments of DNA. Although other methods are available to carry out these functions, they usually require detailed information, often at the DNA sequence level, about the gene or genes of interest. The value of the vector and approach described here is that *no* information need be known to clone and express fragments of open reading frame DNA from among a larger number of fragments.

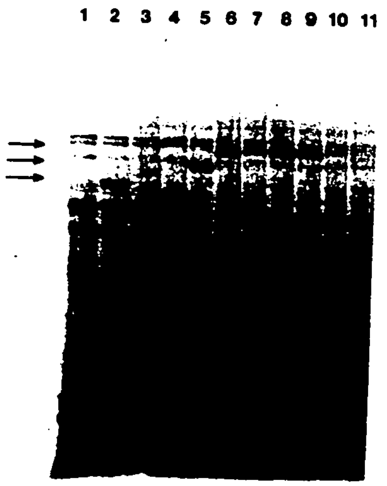

FIG. 3. Gel analysis of hybrid proteins. Proteins from various sources were analyzed by NaDodSO$_4$ gel electrophoresis and the gel was stained with Coomassie blue. Lanes: 1, HB101; 2, strain HB101 was grown overnight at 37°C in the presence of 1 mM isopropyl-β-D-thiogalactoside; 3, LG90; 4, pMR100; 5, pMR200; 6–11, clones 1, 3, and 5–8, respectively (see Table 2). Arrows: lower, wild-type lacZ (lane 2); middle, cI–lacIZ fusion protein (lane 5); upper, large insert-derived fusion proteins (lanes 6–9 and 11).

The data presented have been chosen from a large number of similar experiments to illustrate some important features of the methodology. When shotgun-cloning random fragments of DNA, it is important to size the DNA carefully prior to ligation. Because the frequency of random open reading frames increases in inverse proportion to DNA fragment size, the selection of red colonies will also select for small DNA inserts, even if the average DNA size is quite large (data not shown). In the absence of any information about the DNA to be cloned in this way, we have chosen a DNA insert size of 400–500 bp. The probability that a random piece of DNA of length x is the correct size $(3n - 1)$ and entirely open reading frame $\approx (1/3) \times (0.953)^x \approx 5.5 \times 10^{-4}$ for 400-bp DNA and 1.1×10^{-4} for 500-bp DNA. The probability that a segment of DNA that is derived from a larger piece of open reading frame DNA is of the correct length and entirely open reading frame = the probability that it is the correct length (1/3) × the probability that it is in the correct frame (1/3) × the probability that it has cloned in the correct orientation $(1/2) = 5.5 \times 10^{-2}$. The difference between these numbers suggests that an open reading frame clone from a 400- to 500-bp DNA insert can be considered with some confidence to

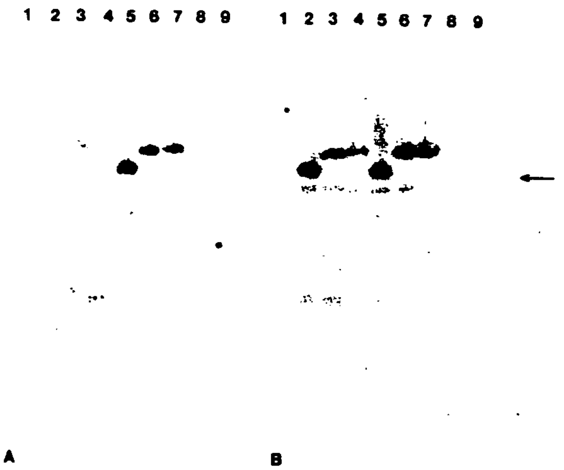

FIG. 4. Blot analysis of hybrid proteins. Proteins from various sources were analyzed by protein blotting and antibody staining with anti-cI antibody (*A*) and anti-β-galactosidase antibody (*B*). Lanes: 1, pMR100; 2, pMR200; 3–9, clones 1 and 3–8, respectively (see Table 2). The arrow indicates the small protein discussed in the text. The additional smaller protein bands present in all lanes (except lane 1) are probably fusion protein degradation products that lack discrete portions from the carboxyl terminus, since these bands are visible with anti-cI antibody.

be a bona fide open reading frame in the DNA from which it was derived. This level of confidence decreases considerably as a function of insert size. Such an argument suggests that genes with small exons and large introns are not generally approachable in this way although sonicated double-stranded cDNA could be cloned in a similar fashion.

The data in Table 1 argue that the strategy enriches significantly for open reading frame DNA segments. On the other hand, some false positives are certainly generated, as shown in Table 2 and Figs. 3 and 4. Some red colonies contain plasmid DNA with no detectable DNA insert and a lacZ polypeptide of a molecular weight identical to the initial cI–lacIZ fusion protein. As indicated above, these colonies are probably due to a small amount of exonuclease activity present during *Sma* I digestion or ligation. This source of red colonies is more problematic when the fraction of bona fide open reading frame colonies is low [e.g., when cloning genomic DNA from higher eukaryotes (unpublished experiments)]. There are also red colonies with proper DNA inserts but no detectable large fusion proteins. We have generated colonies of this phenotype from many sources of eukaryotic DNA (data not shown). They are always associated with the presence of a small lacZ polypeptide of molecular weight similar or identical to the lacIZ-encoded portion of the fusion protein. Because a similar lacZ polypeptide is visible in protein from pMR200 and all of these small polypeptides fail to react with anti-cI antibody, we favor the interpretation that they are due to proteolysis of the larger cI-positive fusion protein. Presumably, this occurs to a variable extent with different fusion proteins. Alternatively, these small proteins from the insert DNA-containing plasmids may be due to translational restarts near the insert–*lacIZ* junction. This matter might be clarified by determination of the DNA sequence of several inserts from plasmids of this type.

Despite these two types of fusion proteins, a substantial fraction of red colonies contain bona fide open reading frame inserts as determined by the nature of the fusion protein they generate (Fig. 4). We expect this general approach to prove useful for a number of purposes. In addition to those described here, many of the fusion proteins generated in this or in a similar way may be antigenic and immunogenic. Large numbers of lac⁺ colonies could be selected on lactose minimal plates and screened wit immunological reagents. Thus, it should be feasible to clon specific genes on this basis. The methodology should also serv as a rapid and convenient way to proceed in the opposite di rection [i.e., from a gene (or a subgene region) to an antibod reagent directed against a portion of the proteins coded for b the starting DNA].

We are grateful to the individuals who supplied reagents. Also, w thank Drs. Donald Winkelmann and L. E. Cannon for advice abou protein blotting and M. A. Osley and L. M. Hereford for helpful dis cussions. This work was supported by National Institutes of Healt Grant HD-08887 (to M.R.). M.R.G. was supported by a predoctora Kessner Fellowship.

Note Added in Proof. The large proteins from clones 5 and 6 react pos itively with antixanthine dehydrogenase antiserum (a gift of Arthu Chovnick), showing that the eukaryotic protein pieces are antigenic.

1. Bassford, P., Beckwith, J., Berman, M., Brickman, E., Casadaban, M., Guarente, L., Saint-Girous, I., Sarthy, A., Schwartz M. & Silhavy, T. (1978) in *The Operon*, eds. Miller, J. H. & Reznikoff, W. S. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 245–262.
2. Guarente, L. & Ptashne, M. (1981) *Proc. Natl. Acad. Sci. USA* 78. 2199–2203.
3. Rose, M., Casadaban, M. J. & Botstein, D. (1981) *Proc. Natl Acad. Sci. USA* 78, 2460–2464.
4. Miller, J. H. (1972) *Experiments in Molecular Genetics* (Col Spring Harbor Laboratory, Cold Spring Harbor, NY).
5. Clewell, D. B. & Helinski, D. R. (1969) *Proc. Natl. Acad. Sci USA* 62, 1159–1166.
6. Ish-Horowicz, D. & Burke, J. F. (1981) *Nucleic Acids Res.* 9 2989–2998.
7. Maniatis, T., Jeffrey, A. & Van de Sande, H. (1975) *Biochemistr* 14, 3787–3794.
8. Dagert, M. & Ehrlich, S. D. (1979) *Gene* 6, 23–28.
9. Laemmli, U. (1970) *Nature (London)* 227, 680–685.
10. Towbin, H., Staehelin, T. & Gordon, J. (1979) *Proc. Natl. Acad. Sci. USA* 76, 4350–4354.
11. Backman, K., Ptashne, M. & Gilbert, W. (1976) *Proc. Natl. Acad Sci. USA* 73, 4174–4178.
12. Guarente, L., Lauer, G., Roberts, T. M. & Ptashne, M. (1980 *Cell* 20, 543–553.
13. Sauer, R. T. (1978) *Nature (London)* 276, 301–302.